

# **FOLLOW THE FLOW: FINE-GRAINED FLOWCHART ATTRIBUTION WITH NEUROSYMBOLIC AGENTS**



Manan Suri, Puneet Mathur, Nedim Lipka,  
Franck Dernoncourt, Ryan A Rossi, Vivek Gupta, Dinesh Manocha

# ATTRIBUTION IN FLOWCHARTS

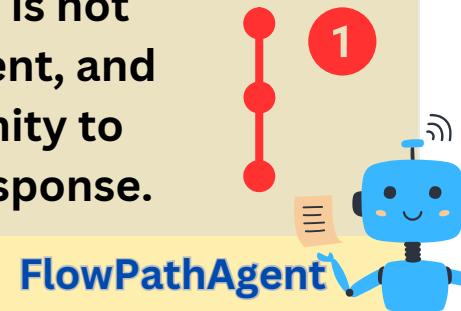
Attribution (represented by with FlowPathAgent ensures logical consistency in flowchart-based reasoning. FlowPathAgent uses a neurosymbolic approach to generate attribution paths ( & ) in the flowchart. This enhances interpretability and reliability in flowchart driven automated decision-making

What is the immediate next step after utilizing prepared items for seeking help, and what decision led to this step?

## Hallucinated Response

The immediate next step is notifying trusted contact of travel plans, and this step was motivated by a positive response to the need to leave the vehicle

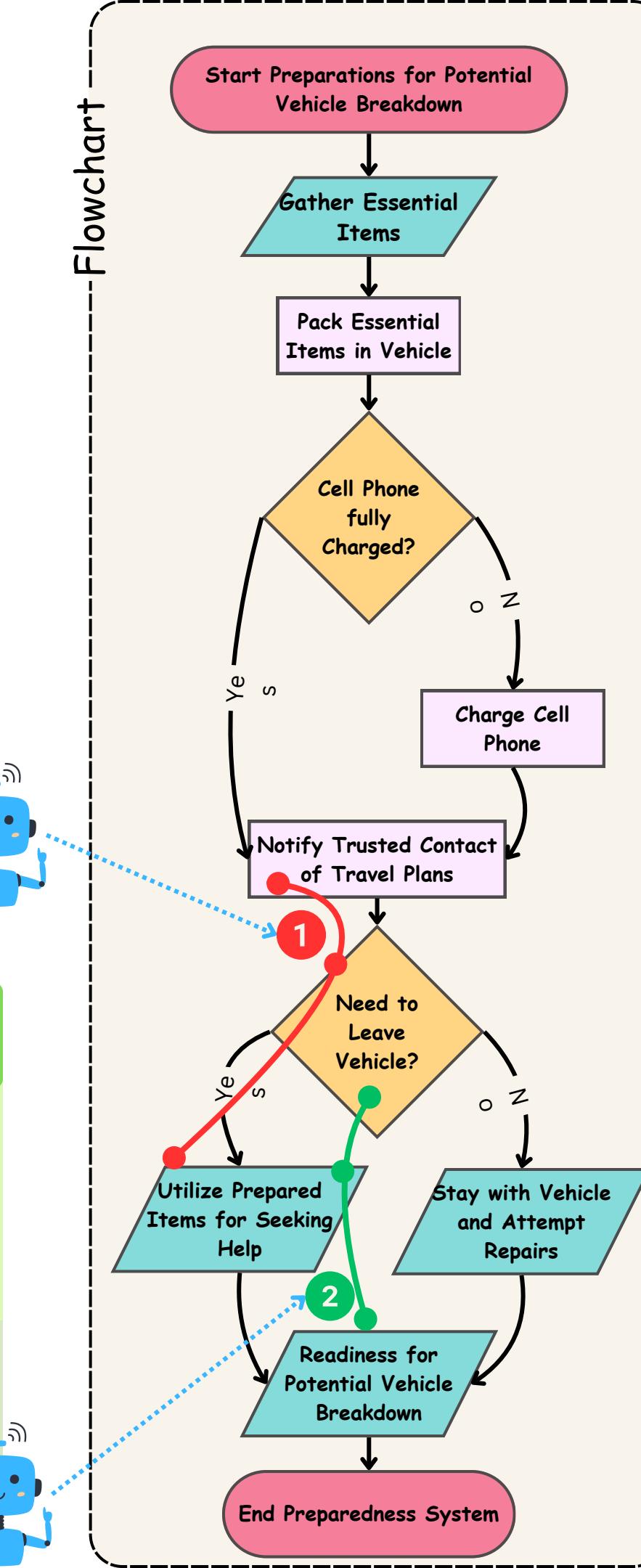
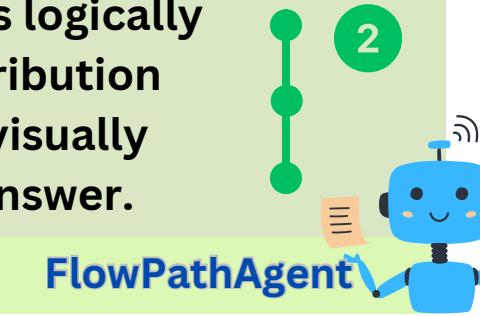
Attributed path is not logically consistent, and offers opportunity to eliminate this response.



## Correct Response

The immediate next step is 'Readiness for Potential Vehicle Breakdown', which follows a 'Yes' decision at the 'Need to Leave Vehicle?' node.

Attributed path is logically consistent. Attribution validates and visually grounds the answer.



# TASK DEFINITION

We formalize fine-grained post-hoc Flowchart Attribution as follows: Given a dataset  $\mathcal{D}$  consisting of a set of flowchart images  $\mathcal{F}$ , each flowchart image  $c_i \in \mathcal{F}, c_i = \mathcal{I}^{w \times h \times 3}$  corresponds to a logical graph representation  $G_i = (V_i, E_i)$ , where  $V_i$  represents the set of nodes and  $E_i$  represents the edges between them. Each node corresponds to a logical operation or directive statement, and the edges represent the flow between these operations. Additionally, the input includes a flowchart-referring statement  $s_i$ , which is a natural language description of a process or action to be grounded in the flowchart image. The underlying goal is to find a path in the image that grounds the statement  $s_i$ . This path may be disjoint, but it should correspond to a set of regions in the flowchart image. The regions are the physical abstraction that corresponds to the logical nodes in the graph. Formally, the task can be represented as a mapping function:

$$F : (c_i, s_i) \mapsto \mathcal{R}_{s_i},$$

where  $F$  maps the flowchart image  $c_i$  and the statement  $s_i$  to a set of regions  $\mathcal{R}_{s_i}$  in the image.  $\mathcal{R}_{s_i} = \{r_{i1}, r_{i2}, \dots, r_{in}\}$  represents the sequence of regions in the image that correspond to a path of logical nodes, and the edges included between consecutive nodes  $v_{i1}, v_{i2}, \dots, v_{in}$  in the graph  $G_i$ , grounding the statement  $s_i$ . The path may be disjoint, but it should satisfy the following criteria:

- 1. Optimality:** The path should be the shortest sequence of regions that ground the statement  $s$ .
- 2. Contextual Alignment:** The path should correspond to the relevant actions and decisions described in  $s$ , matching the flow of the process.
- 3. Exclusivity:** No additional regions outside of  $\mathcal{R}_{s_i}$  are necessary to fully explain the statement  $s$ .

# FLOW EXPLAIN BENCH

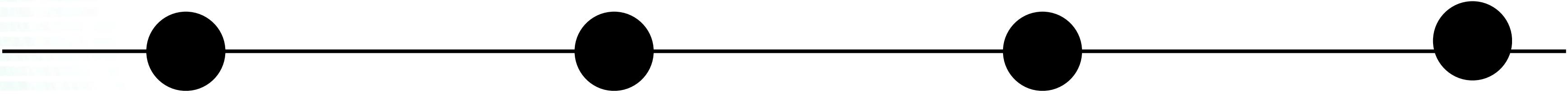
## Data Source

- FlowVQA Dataset

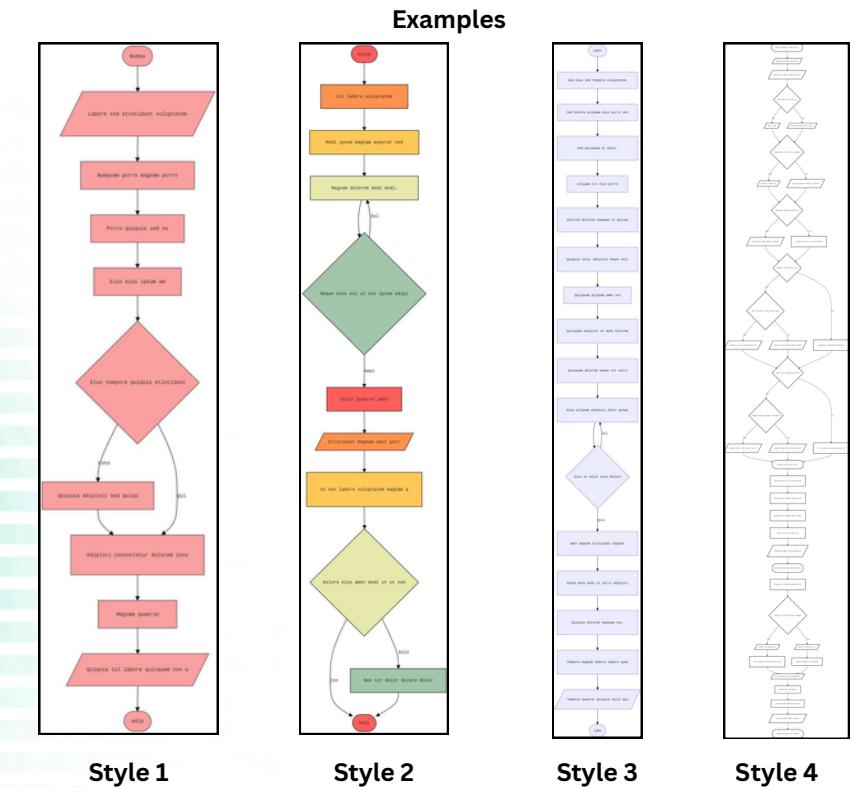
Singh, Shubhankar, et al. "FlowVQA: Mapping multimodal logic in visual question answering with flowcharts." arXiv preprint arXiv:2406.19237 (2024).

## Attribution Annotation

1. Automatic Labeling
2. Human Verification
3. Question Filtering



## Visual Diversity Augmentation



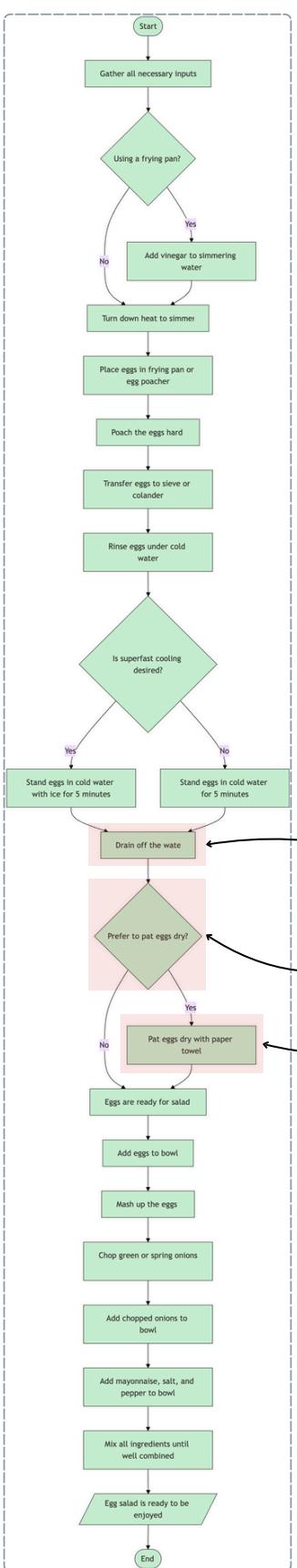
Style 1 Color Options:



## Dataset Composition

|                                   | Code  | Wiki  | Instruct | Overall |
|-----------------------------------|-------|-------|----------|---------|
| <b># of Flowcharts</b>            | 189   | 470   | 294      | 953     |
| <b># of Questions</b>             | 246   | 610   | 382      | 1238    |
| Fact Retrieval                    | 88    | 163   | 102      | 353     |
| Applied Scenario                  | 69    | 128   | 90       | 287     |
| Flow Referential                  | 43    | 128   | 87       | 258     |
| Topological                       | 46    | 191   | 103      | 340     |
| <b>Avg # of Nodes</b>             | 11.85 | 24.49 | 21.59    | 21.08   |
| <b>Max # of Nodes</b>             | 29    | 43    | 44       | 44      |
| <b>Avg Attributed Path Length</b> | 2.59  | 3.21  | 2.88     | 2.99    |
| <b>Max Attributed Path Length</b> | 15    | 35    | 21       | 35      |
| <b>Avg Words (Question)</b>       | 26.99 | 26.12 | 26.56    | 26.43   |
| <b>Avg Words (Answer)</b>         | 8.62  | 8.74  | 9.50     | 8.95    |

# DATASET EXAMPLES

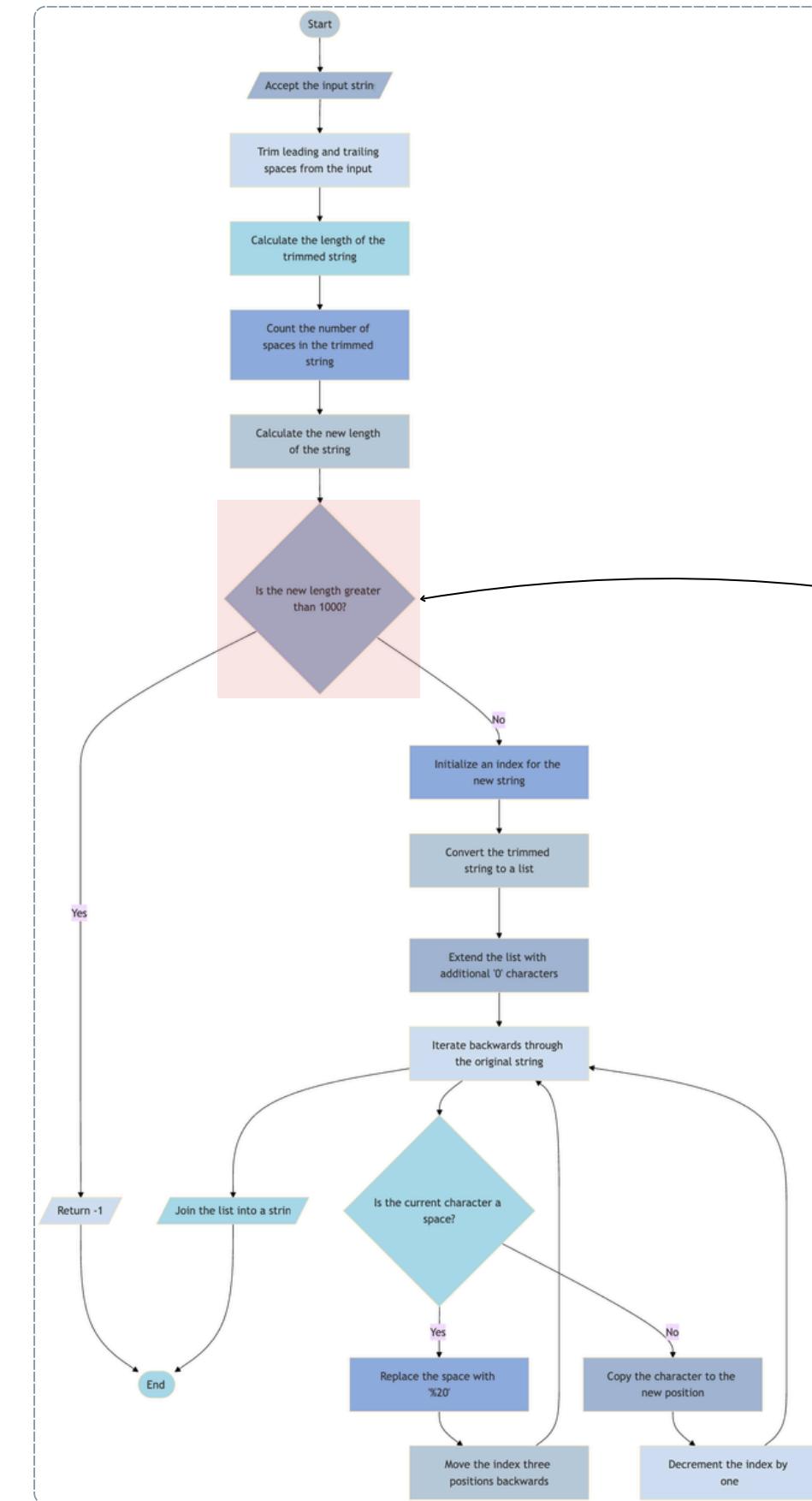


Instruct  
Applied Scenario

**Q:** Emma is making egg salad and prefers her ingredients to be as dry as possible before mixing to avoid a watery salad. After boiling and cooling the eggs, what should her next step be according to the blog post instructions?

**A:** Pat the eggs dry with a paper towel.

Ground Truth  
Attributions



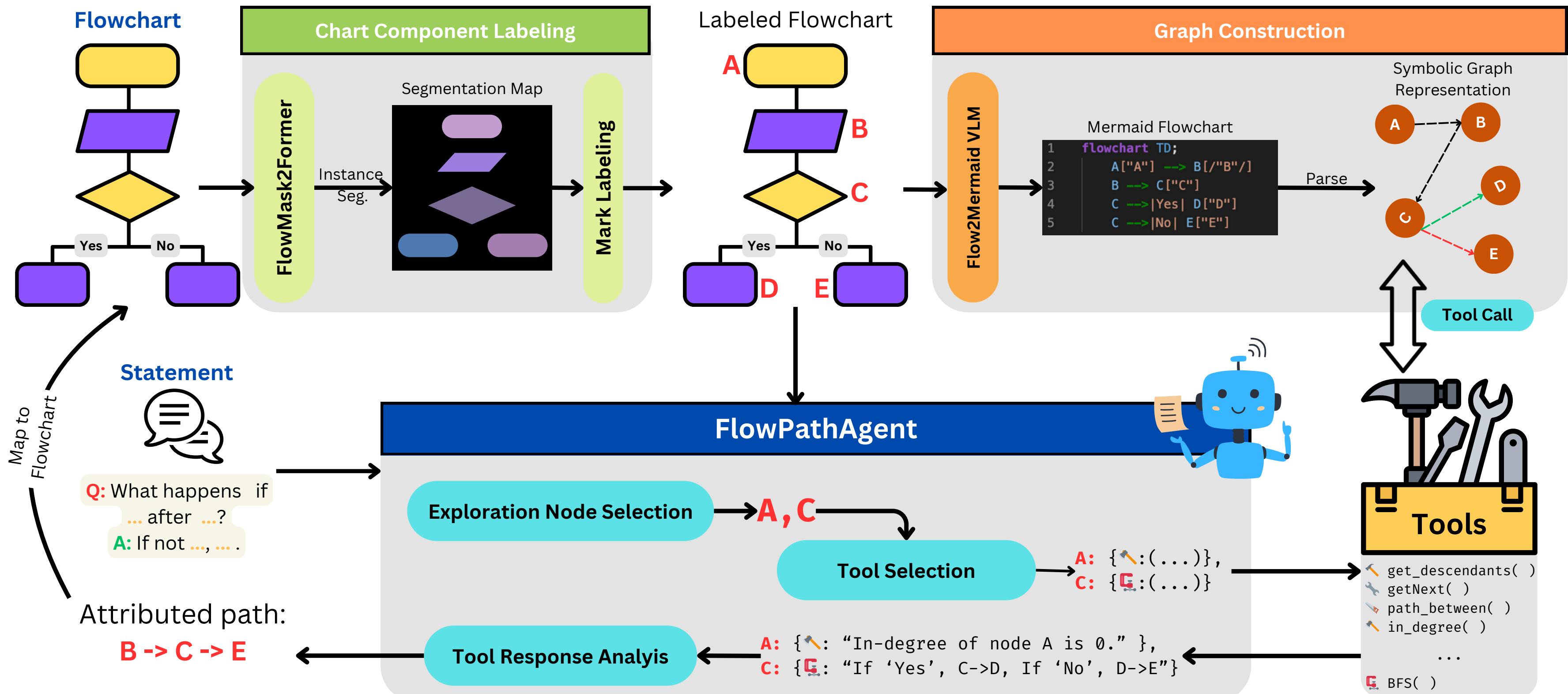
Code  
Fact Retrieval

**Q:** What is the maximum allowed length for the new string after spaces are replaced?

**A:** The new string must not exceed 1000 characters in length.

Ground Truth  
Attribution

# FLOW PATH AGENT

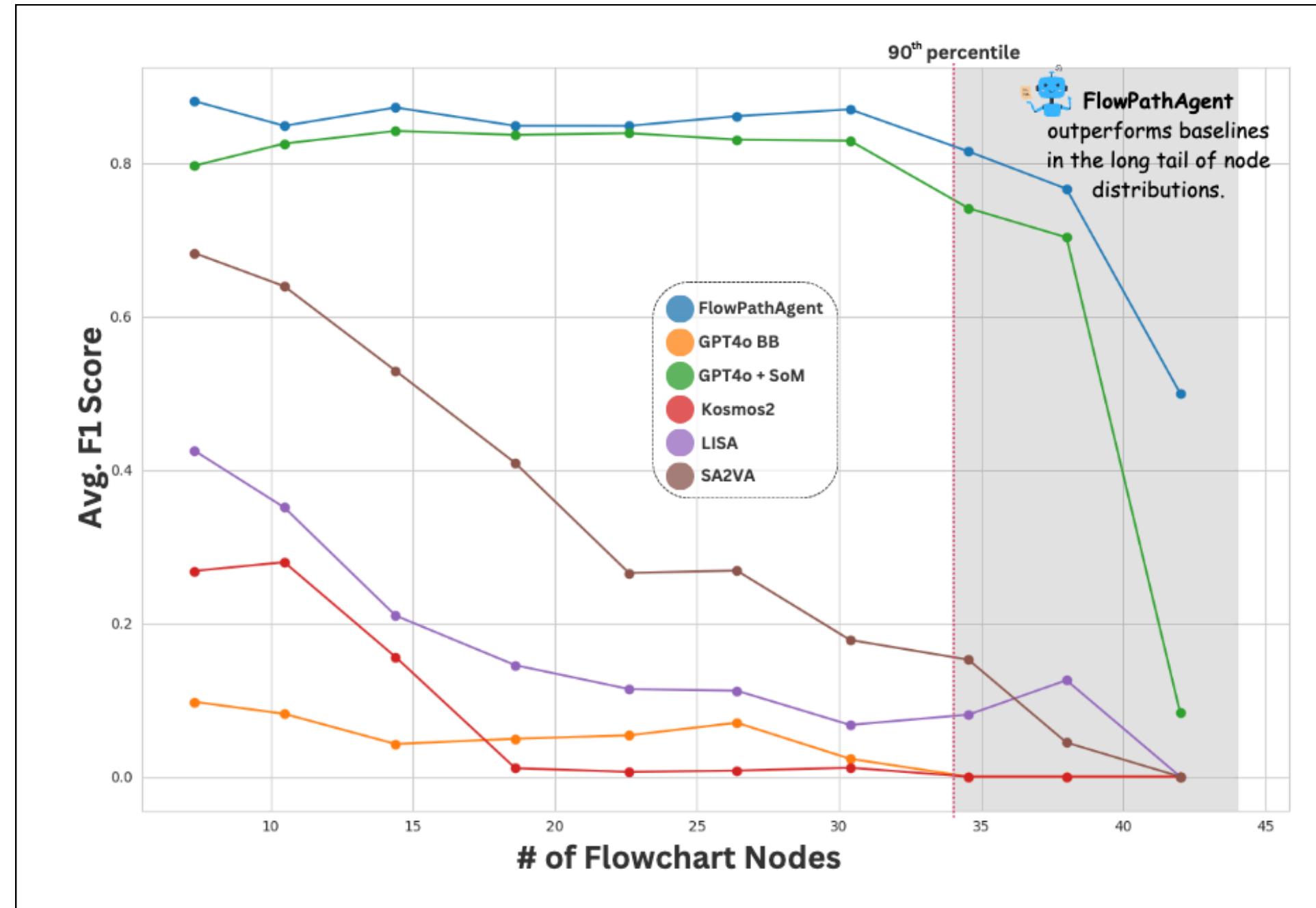


# RESULTS

| Baseline   | Overall   |        |       | FEBench-Code |        |       | FEBench-Wiki |        |       | FEBench-Instruct |        |       |
|--|-----------|--------|-------|--------------|--------|-------|--------------|--------|-------|------------------|--------|-------|
|  | Precision | Recall | F1    | Precision    | Recall | F1    | Precision    | Recall | F1    | Precision        | Recall | F1    |
| Kosmos-2 ( <a href="#">Peng et al., 2023</a> )       | 37.14     | 1.76   | 3.36  | 41.41        | 6.45   | 11.16 | 20.69        | 0.31   | 0.60  | 38.30            | 1.64   | 3.14  |
| LISA ( <a href="#">Lai et al., 2024</a> )            | 18.01     | 14.34  | 15.97 | 35.36        | 19.18  | 24.87 | 14.09        | 11.74  | 12.81 | 18.45            | 16.18  | 17.24 |
| SA2VA ( <a href="#">Yuan et al., 2025</a> )          | 66.36     | 9.88   | 17.20 | 79.35        | 19.34  | 31.10 | 58.47        | 7.40   | 13.14 | 65.99            | 8.82   | 15.56 |
| VisProg ( <a href="#">Gupta and Kembhavi, 2022</a> ) | 45.95     | 0.46   | 0.91  | 46.88        | 2.30   | 4.49  | 0.00         | 0.00   | 0.00  | 25.00            | 00.09  | 0.18  |
| GPT4o Bounding Box                                   | 58.82     | 1.90   | 3.68  | 80.00        | 1.89   | 3.69  | 53.19        | 1.29   | 2.51  | 57.89            | 3.00   | 5.70  |
| GPT4o SoM  | 74.10     | 67.69  | 70.75 | 67.32        | 70.28  | 68.77 | 74.55        | 65.03  | 69.47 | 77.84            | 70.91  | 74.22 |
| <b>FlowPathAgent</b>                                 | 77.19     | 77.21  | 77.20 | 74.18        | 80.62  | 77.27 | 76.29        | 74.21  | 75.23 | 80.28            | 80.19  | 80.23 |

Table 2: Performance comparison of FlowPathAgent with baselines on FlowExplainBench. Best and second-best results have been highlighted.

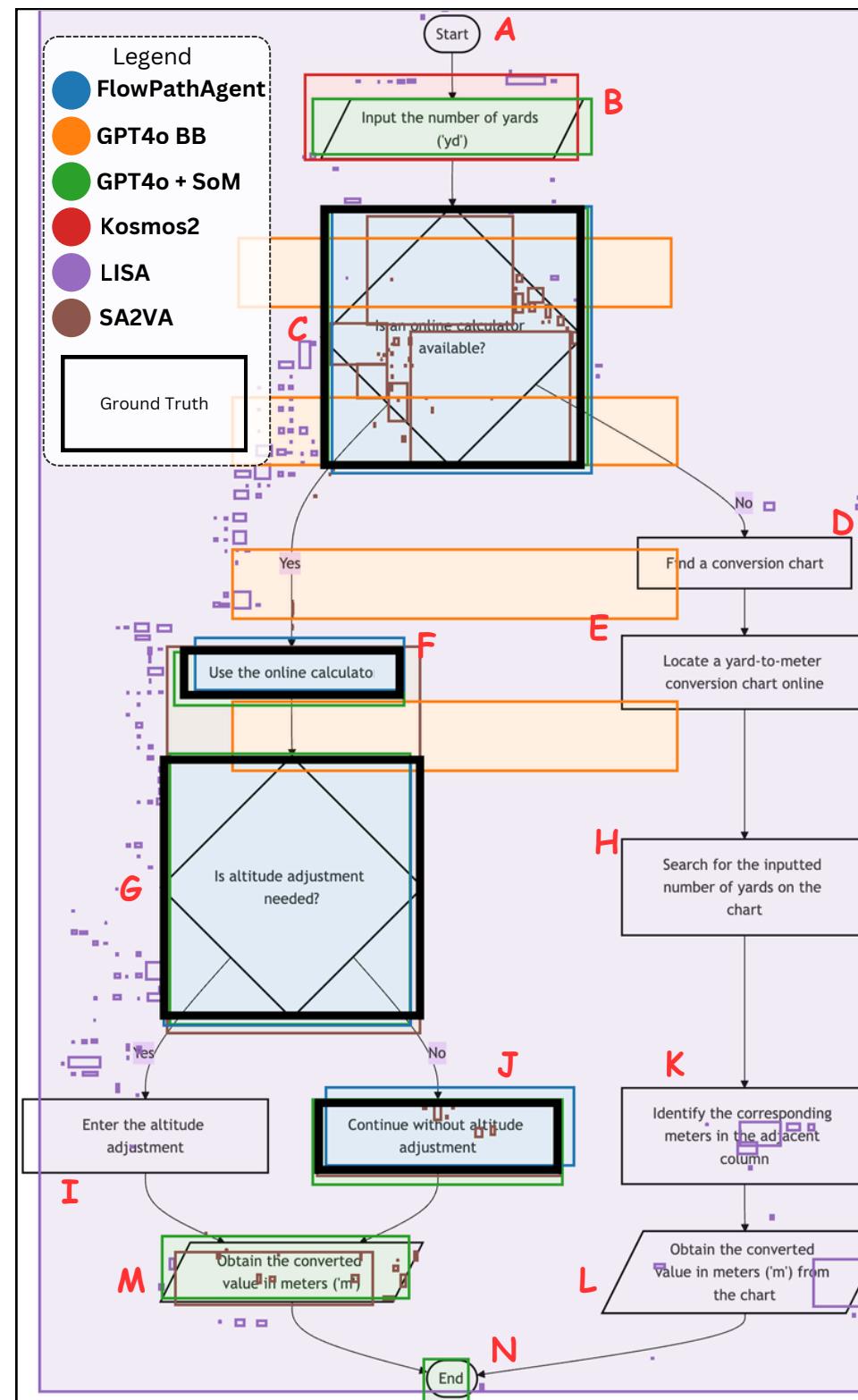
# RESULTS



Performance comparison of FlowPathAgent against baselines demonstrates superior effectiveness across long-tail distribution of node count in flowcharts.

# QUALITATIVE EXAMPLES

## Flowchart (with attributions)



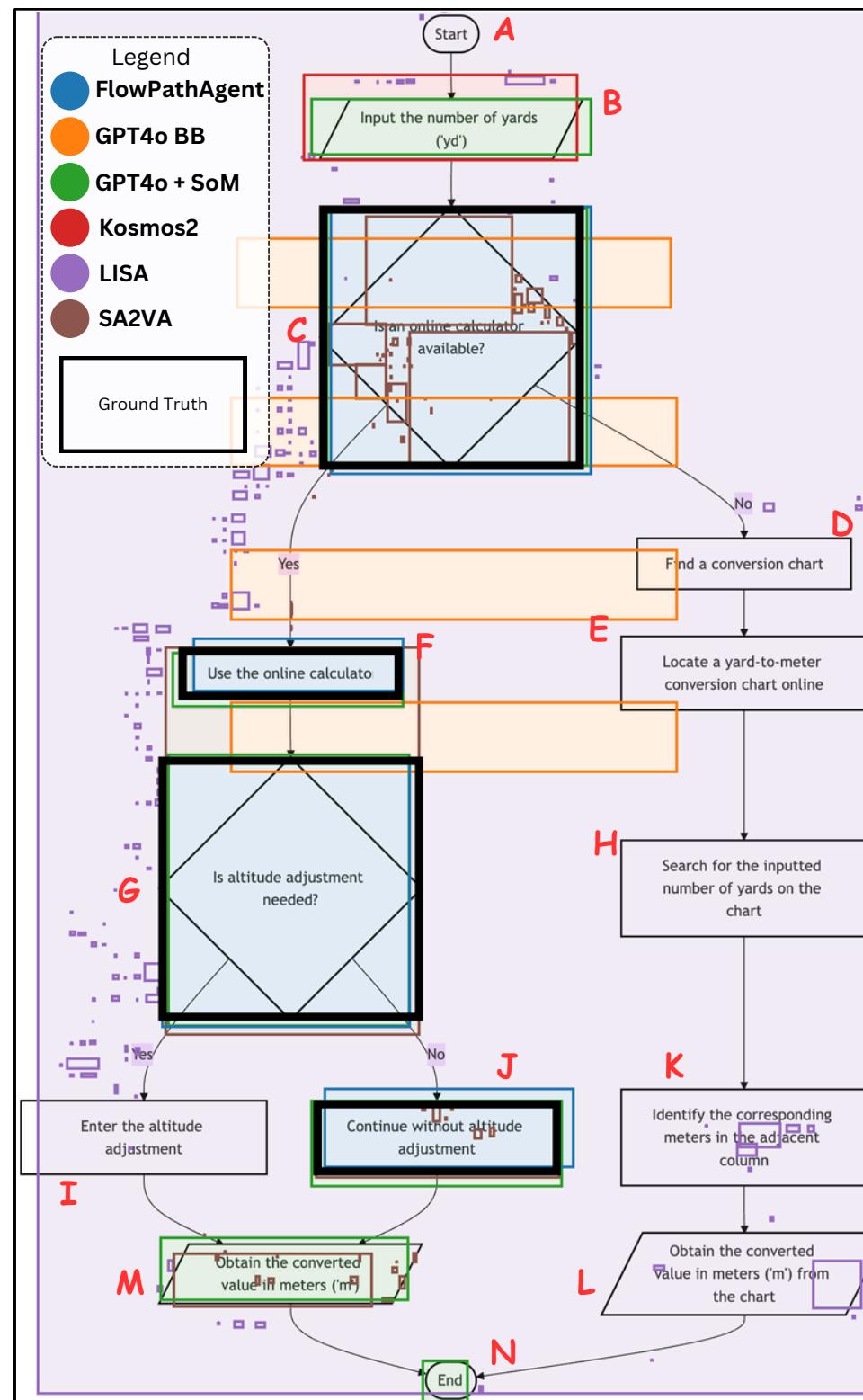
## Statement

**Question:** A landscape architect, Sophia, is working on a garden design that was initially measured in yards. However, the international team she's collaborating with uses the metric system. Sophia opts for an online calculator to convert the measurements but is unsure whether altitude plays a role in this scenario. Should Sophia make any adjustments before finalizing the conversion?

**Answer:** Sophia should continue without altitude adjustment unless the specific context of her work requires it.

# QUALITATIVE EXAMPLES

## Flowchart (with attributions)



### GPT4o + FlowMask2Former Set of Marks X

The Q/A pair involves the process of using an online calculator for conversion and considering altitude adjustment.

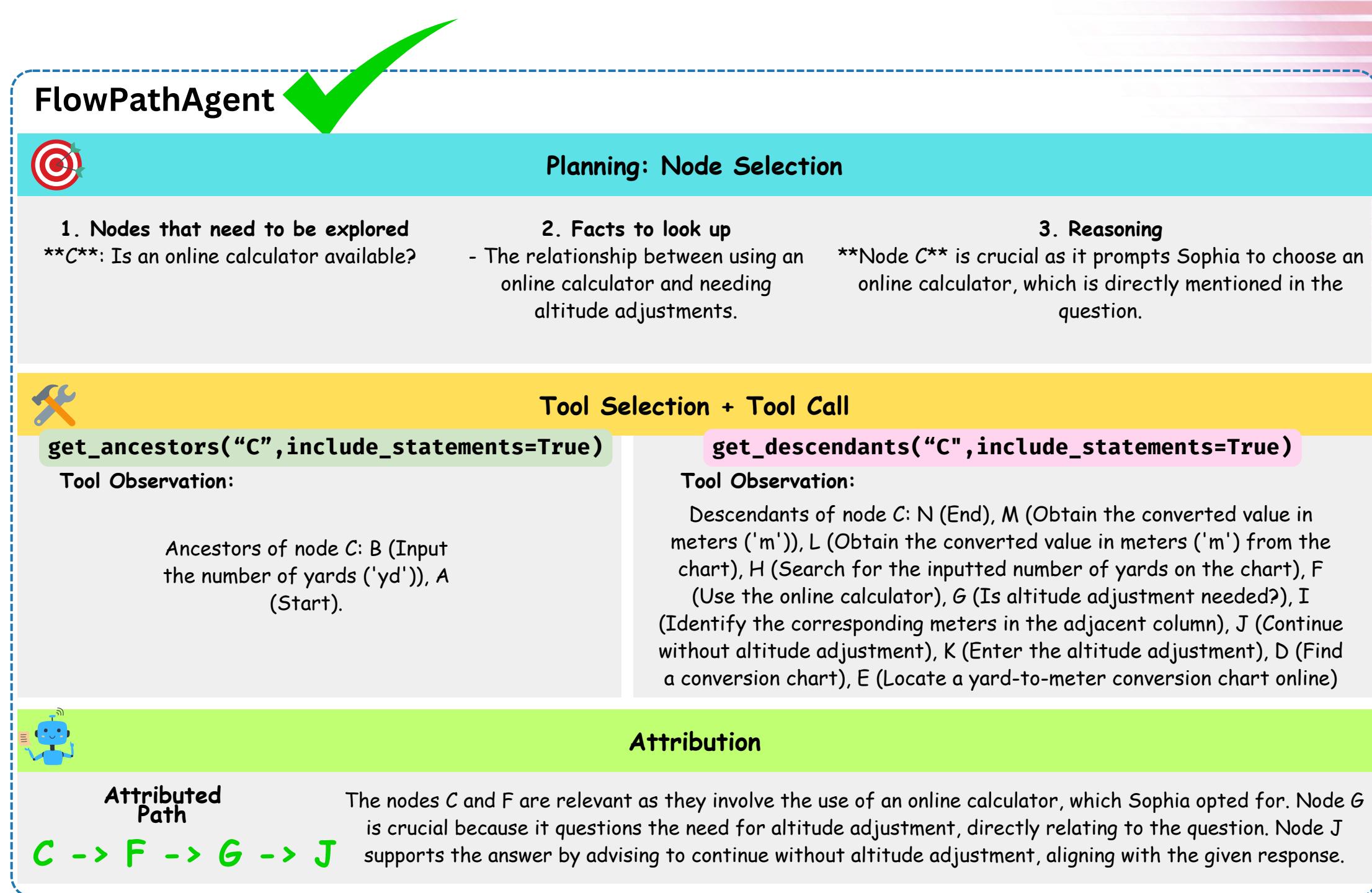
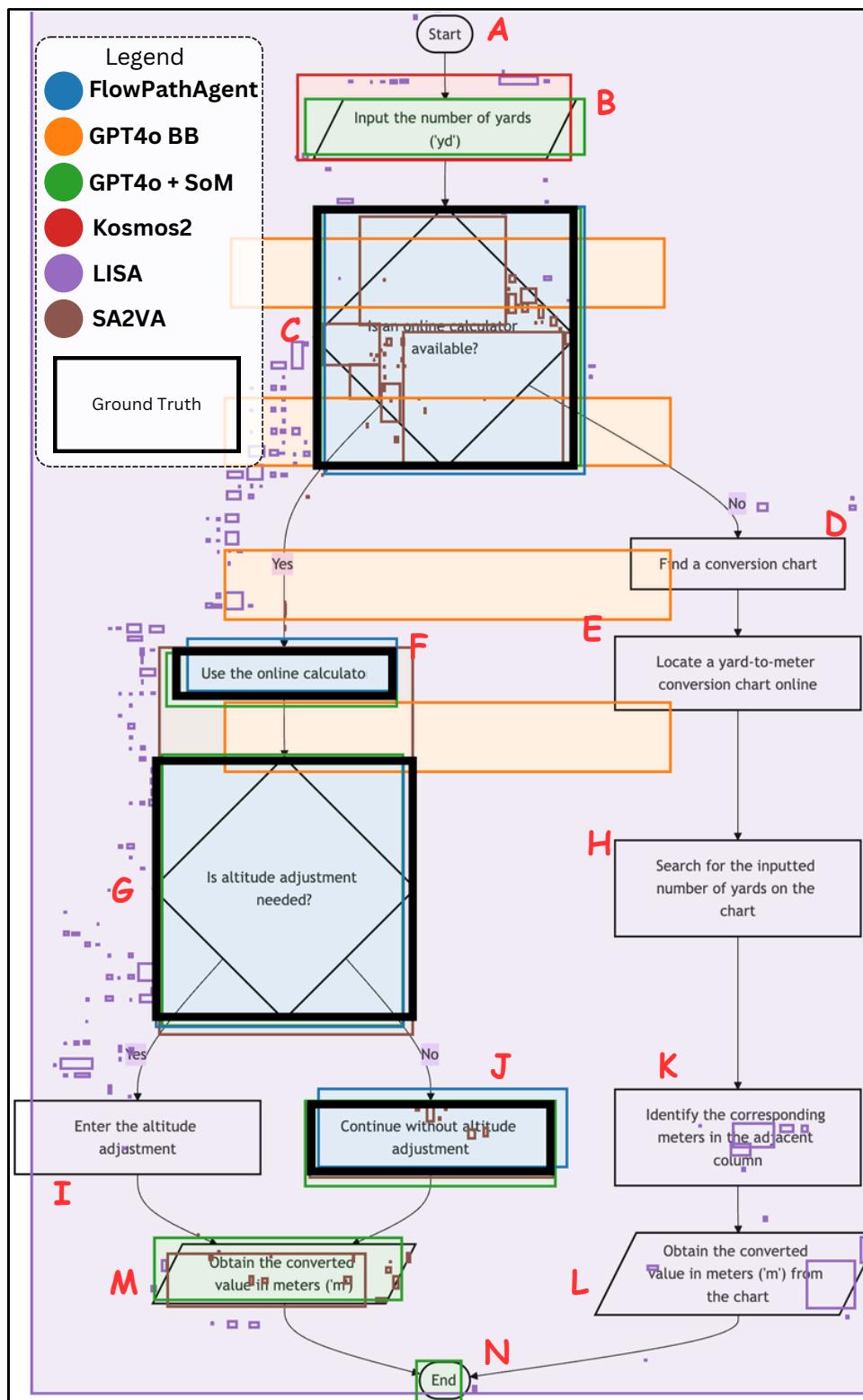
- \*\*Node B\*\* is where the input of yards is initiated.
- \*\*Node C\*\* checks the availability of an online calculator.
- \*\*Node F\*\* involves using the online calculator.
- \*\*Node G\*\* asks if altitude adjustment is needed.
- \*\*Node J\*\* indicates continuing without altitude adjustment.
- \*\*Node M\*\* is where the converted value is obtained.
- \*\*Node N\*\* marks the end of the process.

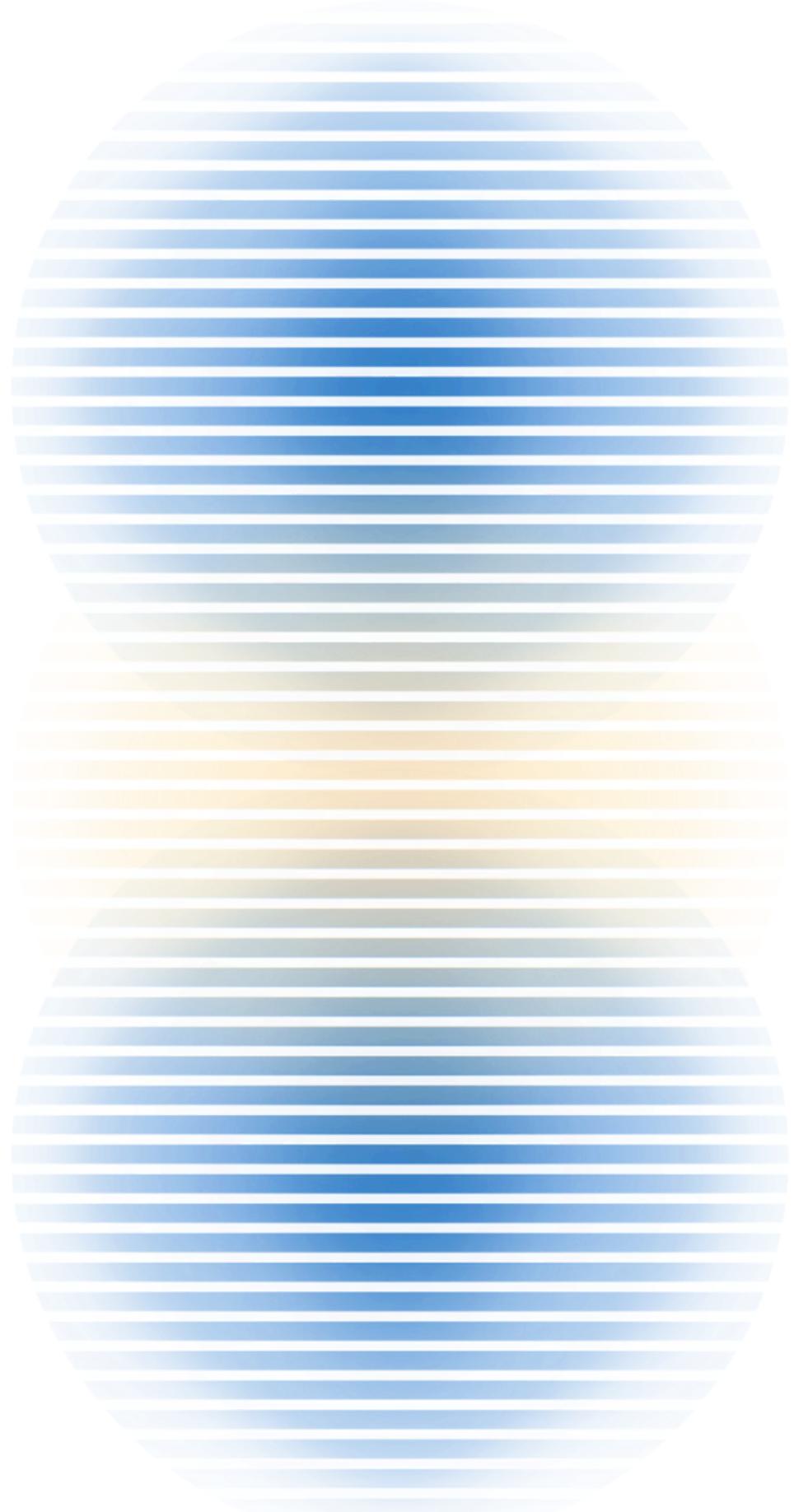
**Attributed Path**

B → C → F → G → J → M → N

# QUALITATIVE EXAMPLES

## Flowchart (with attributions)





# THANK YOU!

Manan Suri  
[manans@umd.edu](mailto:manans@umd.edu)